

# ON SOBOLEV REGULARITY OF MASS TRANSPORT AND TRANSPORTATION INEQUALITIES

<sup>1</sup>Alexander V. Kolesnikov

**ABSTRACT.** We study Sobolev a priori estimates for the optimal transportation  $T = \nabla\Phi$  between probability measures  $\mu = e^{-V} dx$  and  $\nu = e^{-W} dx$  on  $\mathbb{R}^d$ . Assuming uniform convexity of the potential  $W$  we show that  $\int \|D^2\Phi\|_{HS}^2 d\mu$ , where  $\|\cdot\|_{HS}$  is the Hilbert-Schmidt norm, is controlled by the Fisher information of  $\mu$ . In addition, we prove similar estimate for the  $L^p(\mu)$ -norms of  $\|D^2\Phi\|$  and obtain some  $L^p$ -generalizations of the well-known Caffarelli contraction theorem. We establish a connection of our results with the Talagrand transportation inequality. We also prove a corresponding dimension-free version for the relative Fisher information with respect to a Gaussian measure.

**Keywords:** Monge-Kantorovich problem, Monge-Ampère equation, Sobolev a priori estimates, Gaussian measures, log-concave measures, transportation inequalities, log-Sobolev inequality, Lipschitz mappings

## 1. INTRODUCTION

Let  $\mu = e^{-V} dx$  and  $\nu = e^{-W} dx$  be probability measures on  $\mathbb{R}^d$  and let  $T = \nabla\Phi$  be the optimal transportation mapping such that  $\nu$  is the image of  $\mu$  with respect to  $T$ :  $\nu = \mu \circ T^{-1}$ . In what follows we say for brevity that  $T$  sends (pushes forward)  $\mu$  onto  $\nu$ . The corresponding convex potential is denoted by  $\Phi$ . The reader is advised to consult [34] for an account in the optimal transportation theory.

Assuming that  $W$  is uniformly convex ( $D^2W \geq K \cdot \text{Id}$ ,  $K > 0$ ) we prove that

$$(1) \quad \mathcal{I}_\mu := \int |\nabla V|^2 d\mu \geq K \int \|D^2\Phi\|_{HS}^2 d\mu.$$

More generally, we show that for every unit  $e \in \mathbb{R}^d$  and  $p \geq 1$

$$(2) \quad \frac{p+1}{2} \|V_e^2\|_{L^p(\mu)} \geq K \|\Phi_{ee}^2\|_{L^p(\mu)}.$$

These results can be considered as (global, dimension-free) Sobolev a priori estimates for the following Monge-Ampère equation

$$e^{-V} = e^{-W(\nabla\Phi)} \det D^2\Phi.$$

The regularity theory for the Monge-Ampère operator has a quite long history. Many famous scientists contributed to this area. We advise the reader to consult [16] (see also [2], [30], [14], [8], [23], [34]). In particular, some Sobolev a priori estimates for the optimal transportation have been obtained by L. Caffarelli in [6]. The most recent results in this direction are concerned with the Hölder regularity of optimal transportation maps on manifolds (see [32], [25], [10], [19], [13] and the references therein).

---

<sup>1</sup> Moscow State University of Printing Arts, St. Tikhon Orthodox University, and Higher School of Economics (Moscow)

The approach we use here is in a sense probabilistic. The estimates obtained in this paper are 1) dimension-free, 2) global, 3) can be obtained in a constructive way by integration-by-parts and above-tangential formalism. We refer to the works of N. Ivochkina (for instance, [17]) for some similar arguments. In spite of the large amount of results, the only global dimension-free estimate known before was given by the Caffarelli contraction theorem [7]. According to this result every optimal transportation  $T$  sending the standard Gaussian measure onto a log-concave measure  $\nu$  with uniformly convex  $W$  (i.e.  $D^2W \geq K \cdot \text{Id}$  with  $K > 0$ ) is a  $\frac{1}{\sqrt{K}}$ -contraction (i.e.  $\|T\|_{Lip} \leq \frac{1}{\sqrt{K}}$ ).

This contraction theorem has become very popular among probabilists because it gives immediately very nice analytical consequences (for instance, the Bakry-Ledoux theorem, a probabilistic version of the Lévy-Gromov comparison theorem). Some recent generalizations can be found in [21], [18], [33]. Another applications are: log-Sobolev and isoperimetric inequalities. By a recent observation of E. Milman (see [27], [28]), even weaker  $L^p$ -estimates for  $\|D^2\Phi\|$  imply results of this type if the image measure is log-concave.

We note (though it is not aim of this paper) that in this way one can also establish some Sobolev estimates for the third-order derivatives. Our estimates rely on the following (formal) identity:

$$\int V_{x_i}^2 d\mu = \int \langle D^2\Phi \cdot D^2W(\nabla\Phi) \cdot D^2\Phi \cdot e_i, e_i \rangle d\mu + \int \|(D^2\Phi)^{-\frac{1}{2}} D^2\Phi_{x_i} (D^2\Phi)^{-\frac{1}{2}}\|_{\text{HS}}^2 d\mu.$$

In particular, if  $\Phi$  is sufficiently smooth and  $D^2W \geq K \cdot \text{Id}$ ,  $K > 0$ , then this identity implies (1) and the following estimate for the third-order derivatives of  $\Phi$ :

$$\int |\nabla V|^2 d\mu \geq 2\sqrt{K} \int \left[ \sum_{i=1}^d \|D^2\Phi_{x_i}\|_{\text{HS}}^2 \right]^{\frac{1}{2}} d\mu.$$

Another motivation for this study comes from the probability theory. It's worth noting that (1) appears to be very similar to the well-known Talagrand inequality (see [31]), which is a classical representative of the so-called transportation inequalities (see surveys [24], [15]), close relatives of various functional inequalities (concentration, Sobolev, isoperimetric, etc.). Let  $\gamma$  be the standard Gaussian measure. Consider the optimal transportation  $\nabla\Phi$  of  $g \cdot \gamma$  onto  $\gamma$ . Then the following (Talagrand or transportation inequality) holds

$$(3) \quad \text{Ent}_{\gamma} g \geq \frac{1}{2} W_2^2(\gamma, g \cdot \gamma), \quad \text{where}$$

$$\text{Ent}_{\gamma} g = \int g \log g d\gamma, \quad W_2(\gamma, g \cdot \gamma) = \left( \int |x - \nabla\Phi(x)|^2 g d\gamma \right)^{1/2}$$

are the relative entropy and the Kantorovich distance.

We recall that the Talagrand inequality follows from the so-called displacement convexity property of the entropy functional (see [1], [34]). Note in this respect that the energies (Fisher information etc.), unlike entropies, are NOT displacement convex. Nevertheless, in Section 3 we reveal a direct relation of (1) to (3). First we prove the inequality

$$(4) \quad \int (V(x+e) - V(x)) d\mu \geq \frac{K}{2} \int |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 d\mu,$$

where  $e \in \mathbb{R}^d$ . It turns out that (4) can be considered as a version of a generalized Talagrand-type inequality proved in [20]. Then we show that (1) follows from (4) under a natural limiting procedure.

In Section 5 we prove some dimension-free estimates of the type (1). For instance, if  $\mu = g \cdot \gamma$  (with smooth  $g$ ) and  $\nu = \gamma$ , then

$$\begin{aligned} I_\gamma g &= 2\text{Ent}_\gamma g - 2 \int \log \det_2(D^2\Phi - \text{Id}) \, g d\gamma \\ &\quad + \int \|D^2\Phi - \text{Id}\|_{HS}^2 \, g d\gamma + \sum_{k=1}^d \int \text{Tr}[(D^2\Phi)^{-1} D^2\Phi_{x_k}]^2 \, g d\gamma, \end{aligned}$$

where  $I_\gamma g = \int \frac{|\nabla g|^2}{g} d\gamma$  (relative information),  $\det_2(D^2\Phi - \text{Id}) = \det D^2\Phi \cdot \exp(d - \Delta\Phi)$  (the Fredholm-Carleman determinant of  $D^2\Phi - \text{Id}$ ).

We note that all the terms in the right-hand side are non-negative. In particular, this identity implies the following stronger version of the log-Sobolev inequality

$$I_\gamma g \geq 2\text{Ent}_\gamma g - \int \log \det_2(D^2\Phi)^2 \, g \, d\gamma$$

and the following (essentially infinite-dimensional) analog of (1)

$$I_\gamma g \geq \int \|D^2\Phi - \text{Id}\|_{HS}^2 \, g d\gamma.$$

Note that the result stated in this form looks particularly relevant to the Talagrand inequality. See also Remark 5.3 below on uniqueness of the extremals for the classical log-Sobolev inequality. In addition, we prove some dimension-free results for the general log-concave reference measures.

In Section 6 we prove several  $L^p$ -generalizations of the main result. We prove that for every fixed unit vector  $e$  and  $p \geq 1$  one has

$$K \|\Phi_{ee}^2\|_{L^p(\mu)} \leq \|(V_{ee})_+\|_{L^p(\mu)},$$

$$K \|\Phi_{ee}^2\|_{L^p(\mu)} \leq \frac{p+1}{2} \|V_e^2\|_{L^p(\mu)}.$$

We emphasize that all these estimates can be obtained without any use of regularity theory. Instead of it we apply the change of variables formula from [26] and the above-tangential formalism. Note that the contraction theorem follows from these estimates and this is exactly the case when  $p = \infty$ . In addition, in Section 7 we prove the following dimension-free estimate for the operator norm  $\|D^2\Phi\|$

$$K \left( \int \|D^2\Phi\|^{2p} \, d\mu \right)^{\frac{1}{p}} \leq \left( \int \|(D^2V)_+\|^p \, d\mu \right)^{\frac{1}{p}}.$$

Finally, we note that some of our results hold not only for the optimal transportation mappings. For instance, they can be established for the so-called triangular mappings (see [4], [29]). See Section 2 and the forthcoming paper [22].

The author thanks Luigi Ambrosio, Max-Konstantin von Renesse, Michel Ledoux, Emanuel Milman, and Frank Morgan for their interest and stimulating discussions. This work was partially done during the author's visit to the Technische Universität Berlin under the support of the German Academic Exchange Service (DAAD).

## 2. HEURISTIC PROOF

In this section we give a *formal* computation of the main formula of our work. See Sections 3 and 4 for rigorous justifications.

In what follows we denote by  $\mathcal{I}_\mu$  the Fisher information of  $\mu$ :

$$\mathcal{I}_\mu = \int |\nabla V|^2 d\mu$$

and by  $\|A\|_{HS} = \sqrt{\text{Tr}(A \cdot A^T)}$  the Hilbert-Schmidt norm of a matrix  $A$ . For the operator norm we use the standard notation  $\|\cdot\|$ . It will be assumed throughout that  $\mathcal{I}_\mu < \infty$  and that  $\mu$  and  $\nu$  admit the finite second moments. The last condition is automatically satisfied for  $\nu$  if  $D^2W \geq K \cdot \text{Id}$ ,  $K > 0$ .

Let  $T$  be a mapping sending  $\mu$  onto  $\nu$ . We assume that the potentials  $V, W$  are smooth,  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a smooth diffeomorphism satisfying  $\det DT > 0$ . By the change of variables formula

$$e^{-V} = e^{-W(T)} \det DT.$$

Taking the logarithm we obtain

$$(5) \quad V = W(T) - \log \det DT.$$

Choose a unit vector  $e$  and differentiate (5) along  $e$  twice. To this end we apply the following fundamental relation

$$\partial_e \log \det DT = \text{Tr}[DT_e \cdot (DT)^{-1}].$$

Differentiating once again and applying

$$DT_e \cdot (DT)^{-1} + DT \cdot [(DT)^{-1}]_e = 0$$

we get

$$\partial_{ee} \log \det DT = \text{Tr}[DT_{ee} \cdot (DT)^{-1}] - \text{Tr}[DT_e \cdot (DT)^{-1}]^2.$$

Coming back to (5) one gets

$$V_e = \langle \nabla W(T), T_e \rangle - \text{Tr}[DT_e \cdot (DT)^{-1}],$$

(6)

$$V_{ee} = \langle D^2W(T) \cdot T_e, T_e \rangle + \langle \nabla W(T), T_{ee} \rangle - \text{Tr}[DT_{ee} \cdot (DT)^{-1}] + \text{Tr}[DT_e \cdot (DT)^{-1}]^2.$$

Let us integrate (6) over  $\mu$ . Clearly,  $\int V_{ee} d\mu = \int V_e^2 d\mu$ . Let us show that after taking the integral the terms in the middle cancel each other. Indeed, let us denote  $S = T^{-1}$ . One has

(7)

$$\begin{aligned} \int \langle \nabla W(T), T_{ee} \rangle d\mu &= \int \langle \nabla W, T_{ee}(S) \rangle d\nu = \int \text{Tr} D[T_{ee}(S)] d\nu \\ &= \int \text{Tr}[DT_{ee}(S) \cdot DS] d\nu = \int \text{Tr}[DT_{ee}(S) \cdot (DT)^{-1}(S)] d\nu \end{aligned}$$

(8)

$$= \int \text{Tr}[DT_{ee} \cdot (DT)^{-1}] d\mu.$$

Thus we get

$$\int V_e^2 d\mu = \int \langle D^2W(T) \cdot DT \cdot e, DT \cdot e \rangle d\mu + \int \text{Tr}[DT_e \cdot (DT)^{-1}]^2 d\mu.$$

We are interested in two particular cases

1) Optimal transportation mappings.

Optimal transportation mappings have the form  $T = \nabla \Phi$ , where  $\Phi$  is the convex function. In this case one has

$$(9) \quad \int V_{x_i}^2 d\mu = \int \langle D^2 \Phi \cdot D^2 W(\nabla \Phi) \cdot D^2 \Phi \cdot e_i, e_i \rangle d\mu + \int \text{Tr} \left[ (D^2 \Phi)^{-1} D^2 \Phi_{x_i} \right]^2 d\mu.$$

Note that the last integrand is non-negative and admits another representation

$$\text{Tr} \left[ (D^2 \Phi)^{-1} D^2 \Phi_{x_i} \right]^2 = \left\| (D^2 \Phi)^{-\frac{1}{2}} D^2 \Phi_{x_i} (D^2 \Phi)^{-\frac{1}{2}} \right\|_{\text{HS}}^2.$$

Taking the sum over  $i$  we get

$$(10) \quad \mathcal{I}_\mu = \int \text{Tr} \left[ D^2 \Phi \cdot D^2 W(\nabla \Phi) \cdot D^2 \Phi \right] d\mu + \sum_{i=1}^d \int \left\| (D^2 \Phi)^{-\frac{1}{2}} D^2 \Phi_{x_i} (D^2 \Phi)^{-\frac{1}{2}} \right\|_{\text{HS}}^2 d\mu.$$

2) Triangular mappings.

Mappings of this type have the form

$$T = (T_1(x_1), T_2(x_1, x_2), \dots, T_d(x_1, \dots, x_d)),$$

where every  $T_i$  is increasing in  $x_i$ .

It is easy to check that in this case

$$(11) \quad \int V_{x_i}^2 d\mu = \int \langle D^2 W(T) \cdot \partial_{x_i} T, \partial_{x_i} T \rangle d\mu + \sum_{k=i}^d \int \left( \frac{\partial_{x_i x_k} T_k}{\partial_{x_k} T_k} \right)^2 d\mu,$$

$$(12) \quad \mathcal{I}_\mu = \int \text{Tr} \left[ DT \cdot D^2 W(T) \cdot (DT)^* \right] d\mu + \sum_{k=1}^d \int |\nabla \ln \partial_{x_k} T_k|^2 d\mu.$$

### 3. MAIN RESULT

Recall that a function  $W$  is called uniformly convex (uniformly  $K$ -convex) if

$$(13) \quad x \mapsto W(x) - \frac{K}{2} x^2$$

is a convex function for  $K \geq 0$ . For a smooth  $W$  this is equivalent to the condition  $D^2 W \geq K \cdot \text{Id}$ . Everywhere in this paper we deal with the case  $K > 0$  only.

One can introduce in the standard way the weighted Sobolev spaces  $W^{2,p}(\mu)$ . We say that  $f \in L^2(\mu)$  admits a distributional derivative  $f_{x_i} \in L^1(\mu)$  if

$$\int f_{x_i} \xi d\mu = - \int f \xi_{x_i} d\mu + \int f V_{x_i} \xi d\mu$$

for every test function  $\xi$ . Similarly one can define  $W_0^{2,p}(\mu)$  as a completion of the test functions in the corresponding Sobolev norm. It is known that  $W^{2,p}(\mu) = W_0^{2,p}(\mu)$  if  $\mathcal{I}_\mu < \infty$  (see Theorem 5.1 in [11]).

We denote by  $f^+$  the function  $\max\{f, 0\}$  and by  $A^+$  the positive part of a symmetric matrix  $A$  (or zero matrix if  $A \leq 0$ ).

**Theorem 3.1.** *Assume that  $\mathcal{I}_\mu < \infty$ ,  $\mu$  admits the finite second moment, and  $W$  satisfies (13) for some  $K > 0$ . Then  $\Phi \in W^{2,2}(\mu)$  and*

$$(14) \quad \mathcal{I}_\mu \geq K \int \|D^2 \Phi\|_{\text{HS}}^2 d\mu.$$

*Proof. Step 1* ( $V$  and  $W$  are smooth). Assume, in addition, that  $V$  and  $W$  satisfy the following assumptions

- 1)  $V, W \in C^\infty(\mathbb{R}^d)$  and bounded from below
- 2)  $D^2V \leq c \cdot \text{Id}$  for some  $c \in \mathbb{R}$ .

By the Caffarelli's regularity results (see, for instance, Theorem 4.14 of [34] and some justification in [21], Section 4)  $\Phi$  is smooth. Moreover, it follows by the Caffarelli-type arguments from 2) and the uniform convexity of  $W$  that

$$\sup_{x \in \mathbb{R}^d} \|D^2\Phi(x)\| < C$$

for some  $C$  (see, for instance, Theorem 2.2 in [21] and an independent proof in Section 6 below).

Let us show that (9) holds. We take a smooth compactly supported test function  $\xi$ . Multiply (6) by  $\xi$  and integrate over  $\mu$ . Apply integration-by-parts formula (see (7)). One obtains

$$\begin{aligned} (15) \quad & \int V_{x_i}^2 \xi \, d\mu \\ &= \int \langle D^2\Phi \cdot D^2W(\nabla\Phi) \cdot D^2\Phi \cdot e_i, e_i \rangle \xi \, d\mu + \int \|(D^2\Phi)^{-\frac{1}{2}} D^2\Phi_{x_i} (D^2\Phi)^{-\frac{1}{2}}\|_{\text{HS}}^2 \xi \, d\mu \\ &+ \int \partial_{e_i} \xi \cdot V_{x_i} \, d\mu + \int \langle \nabla \xi, (D^2\Phi)^{-1} D^2\Phi_{x_i} \cdot e_i \rangle \, d\mu. \end{aligned}$$

Assume that  $\xi$  has the form  $\xi = \eta(\nabla\Phi)$ , where  $\eta$  is a test function. One has  $\nabla \xi = D^2\Phi \cdot \nabla \eta(\Phi)$ . Using the uniform estimate of  $\|D^2\Phi\|$  one obtains

$$\left| \int \partial_{e_i} \xi \cdot V_{x_i} \, d\mu \right| \leq C \int |\nabla \eta(\nabla\Phi)| |V_{x_i}| \, d\mu \leq C \mathcal{I}_\mu^{\frac{1}{2}} \left( \int |\nabla \eta|^2 \, d\nu \right)^{\frac{1}{2}}.$$

To estimate the last term we integrate by parts

$$\begin{aligned} & \int \langle \nabla \xi, (D^2\Phi)^{-1} D^2\Phi_{x_i} \cdot e_i \rangle \, d\mu = \int \langle \nabla \eta(\nabla\Phi), D^2\Phi_{x_i} \cdot e_i \rangle \, d\mu \\ &= - \int \langle D^2\eta(\nabla\Phi) D^2\Phi \cdot e_i, D^2\Phi \cdot e_i \rangle \, d\mu + \int \langle \nabla \eta(\nabla\Phi), D^2\Phi \cdot e_i \rangle V_{x_i} \, d\mu. \end{aligned}$$

The latter does not exceed

$$\begin{aligned} & C^2 \int \|D^2\eta(\nabla\Phi)\| \, d\mu + C \int |\nabla \eta(\nabla\Phi)| |V_{x_i}| \, d\mu \\ & \leq C^2 \int \|D^2\eta\| \, d\nu + C \mathcal{I}_\mu^{\frac{1}{2}} \left( \int |\nabla \eta|^2 \, d\nu \right)^{\frac{1}{2}}. \end{aligned}$$

Choosing a sequence of test function  $\{\eta_n\}$  such that  $0 \leq \eta_n \leq 1$ ,  $\eta_n \rightarrow 1$  uniformly on every compact set, and  $|\nabla \eta_n|^2 \rightarrow 0$ ,  $\|D^2\eta_n\| \rightarrow 0$  in  $L^1(\nu)$ , we get (9) (hence (14)) for  $V, W$  satisfying 1)-2).

**Step 2** ( $W$  is smooth). Fix a smooth uniform  $K$ -convex function  $W$  and approximate  $\mu$  by smooth measures. We choose a sequence of functions  $\{V_n\}$  such that every  $V_n$  satisfies 1)-2). In addition, we assume that  $\sqrt{\rho_n} \rightarrow \sqrt{\rho}$  in  $W^{1,2}(\mathbb{R}^d)$ , every  $\mu_n = \rho_n \, dx = e^{-V_n} \, dx$  is a probability measure, and  $\sup_n \int |x|^2 d\mu_n < \infty$ .

Note that there exists a subsequence of  $\{\nabla\Phi_n\}$  (denoted again by  $\{\nabla\Phi_n\}$ ) such that  $\nabla\Phi_n \rightarrow \nabla\Phi$  almost everywhere. Indeed, let  $\Psi_n$  be the convex conjugated function to  $\Phi_n$ . Remind that  $\nabla\Phi_n$  and  $\nabla\Psi_n$  are reciprocal. One has

$\sup_n \int |\nabla \Psi_n|^2 d\nu = \sup_n \int |x|^2 d\mu_n < \infty$ . We also require without loss of generality that  $\int \Psi_n d\nu = 0$  (note that  $\Psi_n \in L^2(\nu)$  by the Poncaré inequality for uniform log-concave measures:  $K \int (f - \int f d\nu)^2 d\nu \leq \int |\nabla f|^2 d\nu$ ). Since  $W$  is smooth,  $\sup_n \int_{B_r} |\nabla \Psi_n|^2 dx < \infty$  for every ball  $B_r$ . Using compactness of Sobolev embeddings one can easily show that there exists an a.e. convergent subsequence (denoted again by  $\{\Psi_n\}$ )  $\Psi_n \rightarrow \Psi$ . Since  $\Psi_n$  are convex, one also has  $\nabla \Psi_n \rightarrow \nabla \Psi$  a.e. This implies a.e. convergence of the convex conjugated potentials  $\Phi_n \rightarrow \Phi$  and their gradients  $\nabla \Phi_n \rightarrow \nabla \Phi$ .

Moreover, since

$$\int \|\nabla \Phi_n\|^2 d\mu_n = \int \|x\|^2 d\nu = \int \|\nabla \Phi\|^2 d\mu,$$

one has  $\nabla \Phi_n \cdot \sqrt{\rho_n} \rightarrow \nabla \Phi \cdot \sqrt{\rho}$  strongly in  $L^2(\mathbb{R}^d)$ . In the same way one can check that (again up to a subsequence)  $\partial_{x_i x_j} \Phi_n \sqrt{\rho_n}$  converges weakly in  $L^2(\mathbb{R}^d)$  to some function  $F$ . This implies

$$\int \xi \cdot \partial_{x_i x_j} \Phi_n \sqrt{\rho_n} \sqrt{\rho_n} dx \rightarrow \int \xi \cdot F \sqrt{\rho} dx.$$

In the other hand

$$\int \xi \cdot \partial_{x_i x_j} \Phi_n \rho_n dx = - \int \xi_{x_j} \cdot \partial_{x_i} \Phi_n \rho_n dx - \int \xi \cdot \partial_{x_i} \Phi_n \frac{\partial_{x_j} \rho_n}{\sqrt{\rho_n}} \cdot \sqrt{\rho_n} dx.$$

By the strong convergence  $\nabla \Phi_n \sqrt{\rho_n} \rightarrow \nabla \Phi \sqrt{\rho}$  the latter tends to

$$- \int \xi_{x_j} \cdot \partial_{x_i} \Phi \rho dx - \int \xi \cdot \partial_{x_i} \Phi \partial_{x_j} \rho dx.$$

The relation

$$\int \xi \cdot F \sqrt{\rho} dx = - \int \xi_{x_j} \cdot \partial_{x_i} \Phi \rho dx - \int \xi \cdot \partial_{x_i} \Phi \partial_{x_j} \rho dx$$

implies that the second distributional derivative  $\partial_{x_i x_j} \Phi$  equals to  $F/\sqrt{\rho}$ . Hence  $D^2 \Phi_n \cdot \sqrt{\rho_n} \rightarrow D^2 \Phi \cdot \sqrt{\rho}$  weakly in  $L^2(\mathbb{R}^d)$ . Since the statement holds for the approximating sequence (according to Step 1), by the standard property of the weak convergence

$$\mathcal{I}_\mu = \lim_n \mathcal{I}_{\mu_n} \geq \underline{\lim}_n \int \|D^2 \Phi_n\|^2 d\mu_n \geq \int \|D^2 \Phi\|^2 d\mu.$$

**Step 3.** At the final step we fix  $\mu$  and approximate  $e^{-W}$  by smooth uniformly log-concave probability densities  $e^{-W_n}$  such that  $\int |x|^2 d\nu_n \rightarrow \int |x|^2 d\nu$  and (13) holds for every  $W_n$ . The proof follows the arguments of Step 2. It is even easier because one has to deal with the fixed reference measure  $\mu$ . One obtains that  $\nabla \Phi_n \rightarrow \nabla \Phi$  strongly in  $L^2(\mu)$  and  $D^2 \Phi_n \rightarrow D^2 \Phi$  weakly in  $L^2(\mu)$ . The result follows from the standard properties of the weak convergence.  $\square$

*Remark 3.2. Third-order derivatives.* Note that some global bounds on the third derivatives of  $\Phi$  are also available. Indeed, if  $\Phi$  is sufficiently smooth and (9)

holds, then

$$\begin{aligned} \int V_{x_i}^2 d\mu &\geq K \int \|D^2\Phi \cdot e_i\|^2 d\mu + \int \|(D^2\Phi)^{-\frac{1}{2}} D^2\Phi_{x_i} (D^2\Phi)^{-\frac{1}{2}}\|_{\text{HS}}^2 d\mu \\ &\geq K \int \|D^2\Phi \cdot e_i\|^2 d\mu + \int \frac{\|D^2\Phi_{x_i}\|_{\text{HS}}^2}{\|D^2\Phi\|^2} d\mu, \end{aligned}$$

where  $\|\cdot\|$  is the standard operator norm. Summing over  $i$ , bounding the operator norm by the Hillbert-Schmidt norm, and applying the Cauchy inequality, one obtains

$$\int |\nabla V|^2 d\mu \geq 2\sqrt{K} \int \left[ \sum_{i=1}^d \|D^2\Phi_{x_i}\|_{\text{HS}}^2 \right]^{\frac{1}{2}} d\mu.$$

#### 4. TRANSPORTATION INEQUALITIES

In this section we show that inequality (1) follows from a (generalized) Talagrand inequality.

The following generalization of the Talagrand inequality has been proved in [20]. Let  $f \cdot \nu, g \cdot \nu$  be probability measures,  $\nu = e^{-W} dx$  with  $D^2W \geq K \cdot \text{Id}$ ,  $K > 0$ . Let  $T_f$  ( $T_g$ ) be the optimal transportation mapping pushing forward  $f \cdot \nu$  ( $g \cdot \nu$ ) onto  $\nu$ . Then the following inequality holds

$$(16) \quad \int f \log \frac{f}{g} d\nu \geq \frac{K}{2} \int |T_f - T_g|^2 f d\nu.$$

*Remark 4.1.* The Talagrand inequality in its classical form

$$(17) \quad \int \rho \log \rho d\nu \geq \frac{K}{2} \int |T(x) - x|^2 \rho d\nu$$

holds for any reasonable transportation mapping  $T$  sending  $\rho \cdot \nu$  onto  $\nu$  and satisfying

$$(18) \quad \text{div}(T^{-1}) - d - \log \det D(T^{-1}) \geq 0$$

(this can be checked by the standard transportation arguments, see, for instance, [24]). Then (16) follows from (17) if we set

$$\rho = \frac{f}{g} \circ (T_g^{-1}), \quad T = T_f \circ T_g^{-1}.$$

Note that (18) holds for  $T$  because  $D(T^{-1})$  is a composition of two non-negative matrix (see arguments below in the proof of Theorem 4.3).

Let us apply (16) to  $f(x) = e^{-V(x)+W(x)}$  and  $g(x) = e^{-V(x+e)+W(x)}$  ( $e$  is a fixed vector). Clearly,  $T_f = \nabla\Phi$  is the optimal transportation between  $\mu$  and  $\nu$  and  $T_g = \nabla\Phi(x+e)$ . We obtain

$$\int (V(x+e) - V(x)) d\mu \geq \frac{K}{2} \int |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 d\mu.$$

In order to make the paper self-contained, we give below an independent prove of this result. Then we deduce from it the main result of the paper (inequality (1)).

Recall that every convex function  $\varphi$  admits a.e. the so-called Alexandrov second-order derivative  $D_a^2\varphi$ , which is the absolutely continuous part of its distributional derivative  $D^2\varphi$ .

The following lemma holds trivially for smooth mappings and can be easily checked by approximation arguments.



**Lemma 4.2.** Let  $\varphi : A \rightarrow \mathbb{R}$ ,  $\psi : B \rightarrow \mathbb{R}$  be convex functions on convex sets  $A$ ,  $B$ . Assume that  $\nabla\psi(B) \subset A$ . Then

$$\operatorname{div}(\nabla\varphi \circ \nabla\psi) \geq \operatorname{Tr}[D_a^2\varphi(\nabla\psi) \cdot D_a^2\psi] \, dx \geq 0,$$

where  $\operatorname{div}$  is the distributional derivative.

**Theorem 4.3.** Assume that  $W$  is  $K$ -uniformly convex. Then for every  $e \in \mathbb{R}^d$

$$\int (V(x+e) - V(x)) \, d\mu \geq \frac{K}{2} \int |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 \, d\mu.$$

*Proof.* By a result of R.J. McCann on the change of variables formula (see [26] or [34])

$$e^{-V} = \det_a D^2\Phi \cdot e^{-W(\nabla\Phi)}$$

$\mu$ -a.e. Hence  $V = W(\nabla\Phi) - \log \det_a D^2\Phi$  and

$$\begin{aligned} V(x+e) - V(x) &= W(\nabla\Phi(x+e)) - W(\nabla\Phi(x)) \\ &\quad - \log \left[ (\det_a D^2\Phi(x))^{-1} \cdot \det_a D^2\Phi(x+e) \right]. \end{aligned}$$

By the  $K$ -uniform convexity of  $W$

$$\begin{aligned} W(\nabla\Phi(x+e)) - W(\nabla\Phi(x)) &\geq \langle \nabla\Phi(x+e) - \nabla\Phi(x), \nabla W(\nabla\Phi(x)) \rangle \\ &\quad + \frac{K}{2} |\nabla\Phi(x+e) - \nabla\Phi(x)|^2. \end{aligned}$$

This implies

$$\begin{aligned} \int (V(x+e) - V(x)) \, d\mu &\geq \int \frac{K}{2} |\nabla\Phi(x+e) - \nabla\Phi(x)|^2 \, d\mu \\ &\quad + \int \langle \nabla\Phi(x+e) - \nabla\Phi(x), \nabla W(\nabla\Phi(x)) \rangle \, d\mu \\ &\quad - \int \log \left[ (\det_a D^2\Phi(x))^{-1} \cdot \det_a D^2\Phi(x+e) \right] \, d\mu. \end{aligned}$$

Denote by  $\Psi = \Phi^*$  the convex conjugated function of  $\Phi$ . Using the fact that  $\nabla\Psi$  and  $\nabla\Phi$  are reciprocal we get

$$\begin{aligned} \int \langle \nabla\Phi(x+e) - \nabla\Phi(x), \nabla W(\nabla\Phi(x)) \rangle \, d\mu &= \int \langle \nabla\Phi(\nabla\Psi(x) + e) - x, \nabla W(x) \rangle \, d\nu \\ &= \int \operatorname{div}(\nabla\Phi(\nabla\Psi(x) + e) - x) e^{-W}, \end{aligned}$$

where  $\operatorname{div}(\nabla\Phi(\nabla\Psi(x) + e) - x)$  is the distributional derivative of the vector field  $\nabla\Phi(\nabla\Psi(x) + e) - x$ .

By Lemma 4.2 and the relation  $(D_a^2\Phi(\nabla\Psi))^{-1} = D_a^2\Psi$  which holds  $\nu$ -a.e. (see [26] or [34]), we get

$$\begin{aligned} \int \operatorname{div}(\nabla\Phi(\nabla\Psi(x) + e) - x) e^{-W} &\geq \int (\operatorname{Tr} D_a^2\Phi(\nabla\Psi(x) + e) \cdot D_a^2\Psi(x) - d) \, d\nu \\ &= \int (\operatorname{Tr} D_a^2\Phi(x+e) \cdot (D_a^2\Phi(x))^{-1} - d) \, d\mu. \end{aligned}$$

It remains to note that

$$\operatorname{Tr} D_a^2\Phi(x+e) \cdot (D_a^2\Phi(x))^{-1} - d - \log \left[ \det_a D^2\Phi(x) \cdot (\det_a D^2\Phi(x+e))^{-1} \right] \geq 0.$$

Indeed, if  $A$  and  $B$  are symmetric and non-negative, then

$$\text{Tr}AB - d - \log \det AB = \text{Tr}C - d - \log \det C,$$

where  $C = B^{1/2}AB^{1/2}$  is a symmetric non-negative matrix. It is well-known that  $\text{Tr}C - d - \log \det C > 0$ . Indeed, the latter is equal to  $\sum_i (c_i - 1 - \log c_i) \geq 0$ , where  $c_i$  are eigenvalues of  $C$ . The proof is complete.  $\square$

**Proposition 4.4.** *Inequality (4) implies (1).*

*Proof.* Following the arguments of Theorem 3.1 we see that it is sufficient to establish implication (4)  $\Rightarrow$  (1) for a nice potential  $V$ . By Theorem 4.3

$$\begin{aligned} & \int \frac{V(x+te) + V(x-te) - 2V(x)}{t^2} d\mu \\ & \geq \frac{K}{2t^2} \int \left( \|\nabla\Phi(x+te) - \nabla\Phi(x)\|^2 + \|\nabla\Phi(x-te) - \nabla\Phi(x)\|^2 \right) d\mu. \end{aligned}$$

Thus, without loss of generality we may assume that  $V$  satisfies

$$\frac{\exp(V(x) - V(x+te)) - 1}{t} \rightarrow V_e \text{ in } L^2(\mu), \quad t \rightarrow 0$$

and

$$\lim_{t \rightarrow 0} \frac{1}{t^2} \int (V(x+te) + V(x-te) - 2V(x)) d\mu = \int V_{ee} d\mu$$

for every  $e$ . Extract  $L^2(\mu)$ -weakly convergent subsequences  $\left\{ \frac{\nabla\Phi(x \pm t_n e) - \nabla\Phi(x)}{t_n} \right\}$  (we keep the same index  $n$ ). Note that

$$\begin{aligned} & \int \frac{\nabla\Phi(x + t_n e) - \nabla\Phi(x)}{t_n} \xi d\mu = \int \nabla\Phi(x) \frac{\xi(x - t_n e) - \xi(x)}{t_n} d\mu \\ & + \int \nabla\Phi(x) \xi \frac{\exp(V(x) - V(x - t_n e)) - 1}{t_n} d\mu. \end{aligned}$$

Obviously, the latter tends to

$$- \int \nabla\Phi(x) \xi_e d\mu + \int \nabla\Phi(x) \xi V_e d\mu.$$

Hence

$$\frac{\nabla\Phi(x \pm t_n e) - \nabla\Phi(x)}{t_n} \rightarrow \nabla\Phi_e$$

weakly in  $L^2(\mu)$ . By the properties of the weak convergence.

$$\int V_e^2 d\mu = \int V_{ee} d\mu \geq K \int \|\nabla\Phi_e\|^2 d\mu.$$

Applying this to every  $e_i$  and taking the sum we complete the proof.  $\square$

## 5. DIMENSION-FREE INEQUALITIES

In this section we prove some essentially infinite-dimensional estimates (which do not contain dimension-dependent constants and make sense in the infinite-dimensional case). The results below also hold (with certain modifications) for the triangular mappings.

**5.1. Gaussian case.** We denote by  $\gamma$  the standard Gaussian measure on  $\mathbb{R}^d$ . Let  $\mu = g \cdot \gamma$ ,  $\nu = \gamma$  and  $\nabla \Phi$  be the corresponding optimal transport. According to the result from Section 3

$$\int \left| \frac{\nabla g}{g} - x \right|^2 g d\gamma = \int \|D^2 \Phi\|_{HS}^2 g d\gamma + \sum_{k=1}^d \int \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_{x_k}]^2 g d\gamma.$$

Note that

$$\int \left| \frac{\nabla g}{g} - x \right|^2 g d\gamma = \int \frac{|\nabla g|^2}{g} d\gamma - 2 \int \langle \nabla g, x \rangle d\gamma + \int |x|^2 g d\gamma$$

and

$$\int \|D^2 \Phi\|_{HS}^2 g d\gamma = \int \|D^2 \Phi - \text{Id}\|_{HS}^2 g d\gamma + 2 \int \Delta \Phi g d\gamma - d.$$

Apply integration-by-parts

$$-2 \int \langle \nabla g, x \rangle d\gamma + \int |x|^2 g d\gamma = 2d - \int |x|^2 g d\gamma.$$

By the change of variables formula

$$2 \int \Delta \Phi g d\gamma - d = 2 \int \Delta \Phi g d\gamma - \int |\nabla \Phi|^2 g d\gamma.$$

Consequently

$$\begin{aligned} \int \frac{|\nabla g|^2}{g} d\gamma &= \int \|D^2 \Phi - \text{Id}\|_{HS}^2 g d\gamma + \int (|x|^2 - |\nabla \Phi|^2) g d\gamma + 2 \int (\Delta \Phi - d) g d\gamma \\ &\quad + \sum_{k=1}^d \int \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_{x_k}]^2 g d\gamma. \end{aligned}$$

Taking the logarithm of the change of variables formula we get

$$\log g = \frac{|x|^2}{2} - \frac{|\nabla \Phi|^2}{2} + \log \det D^2 \Phi.$$

Applying this formula we get the heuristic proof of the following statement:

*Every probability measure  $g \cdot \gamma$  with smooth  $g$  and smooth  $\nabla \Phi$  satisfies the following relation*

$$\begin{aligned} (19) \quad I_\gamma g &= 2\text{Ent}_\gamma g - 2 \int \log \det_2(D^2 \Phi - \text{Id}) g d\gamma \\ &\quad + \int \|D^2 \Phi - \text{Id}\|_{HS}^2 g d\gamma + \sum_{k=1}^d \int \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_{x_k}]^2 g d\gamma, \end{aligned}$$

where  $I_\gamma g = \int \frac{|\nabla g|^2}{g} d\gamma$  (relative information),  $\text{Ent}_\gamma g = \int g \log g d\gamma$  (relative entropy),  $\det_2(D^2 \Phi - \text{Id}) = \det D^2 \Phi \cdot \exp(d - \Delta \Phi)$  (the Fredholm-Carleman determinant of  $D^2 \Phi - \text{Id}$ ).

*Remark 5.1.* Since all the terms in the right-hand side are non-negative, this statement implies, in particular, the classical logarithmic Sobolev inequality

$$I_\gamma g \geq 2\text{Ent}_\gamma g$$

and the Gaussian analog of (1)

$$(20) \quad I_\gamma g \geq \int \|D^2 \Phi - \text{Id}\|_{HS}^2 g d\gamma.$$

*Remark 5.2.* Identity (19) holds, for instance, under assumptions:  $g$  is smooth, bounded, strictly positive,  $I_\gamma g < \infty$ , and  $-D^2 \log g \leq c \cdot \text{Id}$ . See Step 1 in the proof of Theorem 3.1.

Inequality (20) follows immediately from Theorem 3.1 under the unique assumption  $I_\gamma g < \infty$ .

*Remark 5.3.* It was pointed out to the author by Michel Ledoux that (19) implies the description of the extremals for the classical log-Sobolev inequality. Indeed, the case of equality in (19) is possible if and only if  $D^2 \Phi = \text{Id}$ , hence  $\nabla \Phi$  is linear and  $g$  has the form  $g = \exp(\langle h, x \rangle - \frac{1}{2} \|h\|^2)$ ,  $h \in \mathbb{R}^d$ . This result has been established by other methods in [9].

**5.2. Log-concave case.** Below we deal with the case  $\mu = g e^{-W} dx$ ,  $\nu = e^{-W} dx$ , where  $W$  is convex. By the above results

$$\int \left| \frac{\nabla g}{g} - \nabla W \right|^2 g d\mu \geq \int \text{Tr} \left[ D^2 \Phi \cdot D^2 W(S) \cdot D^2 \Phi \right] g d\mu.$$

Rewrite the left-hand side

$$\int \left| \frac{\nabla g}{g} - \nabla W \right|^2 g d\mu = \int \frac{|\nabla g|^2}{g} d\mu - 2 \int \langle \nabla g, \nabla W \rangle d\mu + \int |\nabla W|^2 g d\mu.$$

Rewrite the right-hand side

$$\begin{aligned} \int \text{Tr} \left[ D^2 \Phi \cdot D^2 W(\nabla \Phi) \cdot D^2 \Phi \right] g d\mu &= \int \text{Tr} \left[ (D^2 \Phi - \text{Id}) \cdot D^2 W(\nabla \Phi) \cdot (D^2 \Phi - \text{Id}) \right] g d\mu \\ &\quad + 2 \int \text{div}(\nabla W \circ \nabla \Phi) g d\mu - \int \Delta W(\nabla \Phi) g d\mu \\ &= \int \text{Tr} \left[ (D^2 \Phi - \text{Id}) \cdot D^2 W(\nabla \Phi) \cdot (D^2 \Phi - \text{Id}) \right] g d\mu \\ &\quad - 2 \int \langle \nabla g, \nabla W \circ \nabla \Phi \rangle d\mu + 2 \int \langle \nabla W, \nabla W \circ \nabla \Phi \rangle g d\mu - \int \Delta W(\nabla \Phi) g d\mu. \end{aligned}$$

Consequently

$$\begin{aligned} \int \frac{|\nabla g|^2}{g} d\mu + \int |\nabla W|^2 g d\mu &\geq 2 \int \langle \nabla g, \nabla W - \nabla W \circ \nabla \Phi \rangle d\mu + 2 \int \langle \nabla W, \nabla W \circ \nabla \Phi \rangle g d\mu \\ &\quad + \int \text{Tr} \left[ (D^2 \Phi - \text{Id}) \cdot D^2 W(\nabla \Phi) \cdot (D^2 \Phi - \text{Id}) \right] g d\mu - \int \Delta W(\nabla \Phi) g d\mu. \end{aligned}$$

This implies

$$\begin{aligned} \int \frac{|\nabla g|^2}{g} d\mu + \int |\nabla W|^2 g d\mu - 2 \int \langle \nabla W, \nabla W \circ \nabla \Phi \rangle g d\mu + \int |\nabla W \circ \nabla \Phi|^2 g d\mu &\geq \\ \geq 2 \int \langle \nabla g, \nabla W - \nabla W \circ \nabla \Phi \rangle d\mu & \\ + \int \text{Tr} \left[ (D^2 \Phi - \text{Id}) \cdot D^2 W(\nabla \Phi) \cdot (D^2 \Phi - \text{Id}) \right] g d\mu + \int [|\nabla W \circ \nabla \Phi|^2 - \Delta W(\nabla \Phi)] g d\mu. & \end{aligned}$$

Taking into account that

$$\int [|\nabla W \circ \nabla \Phi|^2 - \Delta W(\nabla \Phi)] g d\mu = \int [|\nabla W|^2 - \Delta W] d\mu = 0$$

we get

$$(21) \quad \int \left| \frac{\nabla g}{g} - (\nabla W - \nabla W \circ \nabla \Phi) \right|^2 g d\mu \geq \int \text{Tr} \left[ (D^2 \Phi - \text{Id}) \cdot D^2 W(\nabla \Phi) \cdot (D^2 \Phi - \text{Id}) \right] g d\mu.$$

By the Cauchy inequality

$$(22) \quad 2 \int \frac{|\nabla g|^2}{g} d\mu + 2 \int |\nabla W - \nabla W \circ \nabla \Phi|^2 g d\mu \geq \int \text{Tr} \left[ (D^2 \Phi - \text{Id}) \cdot D^2 W(\nabla \Phi) \cdot (D^2 \Phi - \text{Id}) \right] g d\mu.$$

Thus in order to estimate

$$\int \text{Tr} \left[ (D^2 \Phi - \text{Id}) \cdot D^2 W(\nabla \Phi) \cdot (D^2 \Phi - \text{Id}) \right] g d\mu$$

(or  $\int \|D^2 \Phi - \text{Id}\|_{HS}^2 g d\mu$  for uniformly convex  $W$ ) it is sufficient to get a bound for

$$\int |\nabla W - \nabla W \circ \nabla \Phi|^2 g d\mu.$$

Some estimates of quantities of this type are established in [5]. We give below the proof for the most simple case (the potential has a quadratic-like growth).

**Theorem 5.4.** *Assume that for some  $K > 0$*

$$W(x) - \langle \nabla W(y), x - y \rangle - W(y) \geq \frac{K}{2} |\nabla W(x) - \nabla W(y)|^2$$

and  $D^2 W \geq K \cdot \text{Id}$ . Then

$$\frac{K}{2} \int \|D^2 \Phi - \text{Id}\|_{HS}^2 g d\mu \leq \frac{2}{K} \int g \log g d\mu + \int \frac{|\nabla g|^2}{g} d\mu.$$

In particular, the estimate holds for some  $K > 0$  if  $C_1 \cdot \text{Id} \leq D^2 W \leq C_2 \cdot \text{Id}$ .

*Proof.* The result follows from Theorem 3.1, the above computations, and the estimate below. The proof of the result can be easily reduced to the case of smooth  $g$  and  $T$  (see the proof of Theorem 3.1). By the change of variables formula for  $T = (\nabla \Phi)^{-1} = \nabla \Phi^*$  one has

$$\log g(T) - W(T) + \log \det DT = -W(x).$$

Rewrite it in the following way

$$\log g(T) = W(T) - \langle \nabla W(x), T(x) - x \rangle - W(x) + \left[ \langle \nabla W(x), T(x) - x \rangle - \log \det DT \right].$$

Note that

$$\int \left[ \langle \nabla W(x), T(x) - x \rangle - \log \det DT \right] d\mu = \left[ \text{Tr} DT - d - \log \det DT \right] d\mu \geq 0.$$

Hence

$$\int \log g(T) d\mu \geq \int \left[ W(T) - \langle \nabla W(x), T(x) - x \rangle - W(x) \right] d\mu.$$

By the change of variables

$$\int g \log g d\mu \geq \int \left[ W(x) - \langle \nabla W(\nabla \Phi(x)), x - \nabla \Phi \rangle - W(\nabla \Phi) \right] g(x) d\mu.$$

This inequality, (22), and the assumptions of the Theorem imply the result.  $\square$

## 6. $L^p$ -ESTIMATES AND THE CAFFARELLI'S THEOREM

We generalize below the results of the previous sections and prove some corresponding  $L^p$ -estimates. As a particular case we get the contraction result of Caffarelli. Note that some dimension-free  $L^p$ -generalizations of the Talagrand transportation inequality have been obtained in [5]. In particular, it was shown in [5] that  $\|\nabla\Phi\|_{L^{2p}(\mu)}$  is controlled by  $\int g|\log g|^p d\mu$ ,  $p \geq 1$  for any  $\mu$  satisfying a log-Sobolev inequality.

The proof of the result below follows the arguments of Theorem 4.3. That is why we omit the details and just give a short outline of the proof.

**Theorem 6.1.** *Assume that  $D^2W \geq K \cdot \text{Id}$ . Then for every unit  $e$ ,  $p \geq 0$ , and  $r = \frac{p+2}{2}$  one has*

$$\begin{aligned} K\|\Phi_{ee}^2\|_{L^r(\mu)} &\leq \|(V_{ee})_+\|_{L^r(\mu)}, \\ K\|\Phi_{ee}^2\|_{L^r(\mu)} &\leq \frac{p+4}{4}\|V_e^2\|_{L^r(\mu)}. \end{aligned}$$

*Proof.* Fix unit vector  $e$ , apply the change of variables formula and the uniform convexity of  $W$

$$\begin{aligned} V(x+te) - V(x) &\geq \langle \nabla\Phi(x+te) - \nabla\Phi(x), \nabla W(\nabla\Phi(x)) \rangle \\ &\quad + \frac{K}{2} |\nabla\Phi(x+te) - \nabla\Phi(x)|^2 - \log \left[ (\det_a D^2\Phi(x))^{-1} \cdot \det_a D^2\Phi(x+te) \right]. \end{aligned}$$

Multiply this identity by  $(\delta_{te}\Phi)^p$ , where  $p \geq 0$  and

$$\delta_{te}\Phi = \Phi(x+te) + \Phi(x-te) - 2\Phi(x)$$

and integrate over  $\mu$ . Integrating by parts we get

$$\begin{aligned} &\int \langle \nabla\Phi(x+te) - \nabla\Phi(x), \nabla W(\nabla\Phi(x)) \rangle (\delta_{te}\Phi)^p d\mu \\ &= \int \langle \nabla\Phi(x+te) \circ (\nabla\Psi) - x, \nabla W(x) \rangle (\delta_{te}\Phi)^p \circ (\nabla\Psi) d\nu \\ &\geq \int \left( \text{Tr} [D_a^2\Phi(x+te) \cdot (D_a^2\Phi)^{-1}] \circ (\nabla\Psi) - d \right) (\delta_{te}\Phi)^p \circ (\nabla\Psi) d\nu \\ &\quad + p \int \left\langle \nabla\Phi(x+te) \circ (\nabla\Psi) - x, (D^2\Psi) \nabla\delta_{te}\Phi \circ (\nabla\Psi) \right\rangle (\delta_{te}\Phi)^{p-1} \circ (\nabla\Psi) d\nu. \end{aligned}$$

Applying the inequality  $\text{Tr}A - d - \log \det A \geq 0$  which is valid for compositions of symmetric positive matrices we get

$$\begin{aligned} \int (V(x+te) - V(x)) (\delta_{te}\Phi)^p d\mu &\geq \frac{K}{2} \int |\nabla\Phi(x+te) - \nabla\Phi(x)|^2 (\delta_{te}\Phi)^p d\mu \\ &\quad + p \int \left\langle \nabla\Phi(x+te) - \nabla\Phi(x), (D^2\Psi) \circ \nabla\Phi(x) \nabla\delta_{te}\Phi \right\rangle (\delta_{te}\Phi)^{p-1} d\mu. \end{aligned}$$

Applying the same inequality to  $-te$  and taking the sum we get

$$\begin{aligned} &\int (V(x+te) + V(x-te) - 2V(x)) (\delta_{te}\Phi)^p d\mu \\ &\geq \frac{K}{2} \int |\nabla\Phi(x+te) - \nabla\Phi(x)|^2 (\delta_{te}\Phi)^p d\mu + \frac{K}{2} \int |\nabla\Phi(x-te) - \nabla\Phi(x)|^2 (\delta_{te}\Phi)^p d\mu \\ &\quad + p \int \left\langle \nabla\delta_{te}\Phi, (D_a^2\Phi)^{-1} \nabla\delta_{te}\Phi \right\rangle (\delta_{te}\Phi)^{p-1} d\mu. \end{aligned}$$

Note that the last term is non-negative. Dividing by  $t^{2p}$  and passing to the limit we obtain

$$(23) \quad \int V_{ee} \Phi_{ee}^p d\mu \geq K \int \|D^2\Phi \cdot e\|^2 \Phi_{ee}^p d\mu + p \int \langle (D^2\Phi)^{-1} \nabla \Phi_{ee}, \nabla \Phi_{ee} \rangle \Phi_{ee}^{p-1} d\mu.$$

For the proof of the first part we note that

$$\int V_{ee} \Phi_{ee}^p d\mu \geq K \int \Phi_{ee}^{p+2} d\mu.$$

Applying the Hölder inequality one gets

$$\|(V_{ee})_+\|_{L^{(p+2)/2}(\mu)} \|\Phi_{ee}^p\|_{L^{(p+2)/p}(\mu)} \geq \int V_{ee} \Phi_{ee}^p d\mu.$$

This readily implies the result.

To prove the second part we integrate by parts the left-hand side

$$\begin{aligned} \int V_{ee} \Phi_{ee}^p d\mu &= -p \int V_e \Phi_{eee} \Phi_{ee}^{p-1} d\mu + \int V_e^2 \Phi_{ee}^p d\mu \\ &= -p \int \langle \nabla \Phi_{ee}, V_e \cdot e \rangle \Phi_{ee}^{p-1} d\mu + \int V_e^2 \Phi_{ee}^p d\mu. \end{aligned}$$

By the Cauchy inequality the latter does not exceed

$$p \int \langle (D^2\Phi)^{-1} \nabla \Phi_{ee}, \nabla \Phi_{ee} \rangle \Phi_{ee}^{p-1} d\mu + \frac{p}{4} \int V_e^2 \langle (D^2\Phi)e, e \rangle \Phi_{ee}^{p-1} d\mu + \int V_e^2 \Phi_{ee}^p d\mu.$$

Inequality (23) implies

$$\frac{p+4}{4} \int V_e^2 \Phi_{ee}^p d\mu \geq K \int |\nabla \Phi_e|^2 \Phi_{ee}^p d\mu \geq K \int \Phi_{ee}^{p+2} d\mu.$$

The rest of the proof is the same as in the first part.  $\square$

**Corollary 6.2.** *In the limit  $p \rightarrow \infty$  we obtain the contraction theorem of Caffarelli*

$$K \|\Phi_{ee}\|_{L^\infty(\mu)}^2 \leq \|(V_{ee})_+\|_{L^\infty(\mu)}.$$

## 7. OPERATOR NORM ESTIMATES

This section gives a partial answer to the question asked to the author by Emanuel Milman. Is it possible to estimate effectively (say, without dimension dependence) the operator norm of  $D^2\Phi$ ? Estimates of this type would have interesting consequences for Sobolev-type inequalities of log-concave measures.

Since the operator norm is controlled by the Hilbert-Schmidt norm, the previous results imply trivially the following estimate

$$\mathcal{I}_\mu \geq K \int \|D^2\Phi\|_{HS}^2 d\mu \geq K \int \|D^2\Phi\|^2 d\mu.$$

We emphasize, however, that for many problems the assumption  $\mathcal{I}_\mu < \infty$  is too strong and leads to dimension dependent results.

The main aim of this section is to show that for the uniformly log-concave  $\nu$

$$\int \|(D^2V)^+\| d\mu \geq K \int \|D^2\Phi\|^2 d\mu.$$

**Lemma 7.1.** *Assume that  $\Phi$  is smooth. Then for every smooth vector field  $v$  and every nonnegative test function  $\eta$  the following inequality holds*

$$\begin{aligned} \int \langle D^2 V v, v \rangle \eta \, d\mu &\geq K \int \|D^2 \Phi \cdot v\|^2 \eta \, d\mu + \int \langle (D^2 \Phi)_v \cdot v, (D^2 \Phi)^{-1} \nabla \eta \rangle \, d\mu \\ &\quad + 2 \int \text{Tr}((D^2 \Phi)_v \cdot Dv \cdot (D^2 \Phi)^{-1}) \eta \, d\mu + \int \text{Tr}[(D^2 \Phi)^{-1} (D^2 \Phi)_v]^2 \eta \, d\mu. \end{aligned}$$

*Proof.* It follows from the change of variables formula  $V = W(\nabla \Phi) - \log \det D^2 \Phi$  that

$$\begin{aligned} V(x + tv) - V(x) &= W(\nabla \Phi(x + tv)) - W(\nabla \Phi(x)) \\ &\quad - \log[(\det D^2 \Phi(x))^{-1} \cdot \det D^2 \Phi(x + tv)]. \end{aligned}$$

By the  $K$ -uniform convexity of  $W$

$$\begin{aligned} W(\nabla \Phi(x + tv)) - W(\nabla \Phi(x)) &\geq \langle \nabla \Phi(x + tv) - \nabla \Phi(x), \nabla W(\nabla \Phi(x)) \rangle \\ &\quad + \frac{K}{2} |\nabla \Phi(x + tv) - \nabla \Phi(x)|^2. \end{aligned}$$

This implies

$$\begin{aligned} \int (V(x + tv) - V(x)) \eta \, d\mu &\geq \int \frac{K}{2} |\nabla \Phi(x + tv) - \nabla \Phi(x)|^2 \eta \, d\mu \\ &\quad + \int \langle \nabla \Phi(x + tv) - \nabla \Phi(x), \nabla W(\nabla \Phi(x)) \rangle \eta \, d\mu \\ &\quad - \int \log[(\det D^2 \Phi(x))^{-1} \cdot \det D^2 \Phi(x + tv)] \eta \, d\mu. \end{aligned}$$

Denote by  $\Psi = \Phi^*$  the convex conjugated function of  $\Phi$ . Using the fact that  $\nabla \Psi$  and  $\nabla \Phi$  are reciprocal we get

$$\begin{aligned} &\int \langle \nabla \Phi(x + tv) - \nabla \Phi(x), \nabla W(\nabla \Phi(x)) \rangle \eta \, d\mu \\ &= \int \langle \nabla \Phi(x + tv) \circ (\nabla \Psi) - x, \nabla W(x) \rangle \eta(\nabla \Psi) \, d\nu \\ &= \int \text{div}(\nabla \Phi(x + tv) \circ (\nabla \Psi) - x) \eta(\nabla \Psi) e^{-W} \, dx \\ &\quad + \int \langle \nabla \Phi(x + tv) \circ (\nabla \Psi) - x, D^2 \Psi \cdot \nabla \eta(\nabla \Psi) \rangle \, d\nu. \end{aligned}$$

By the relation  $(D^2 \Phi(\nabla \Psi))^{-1} = D^2 \Psi$  we get

$$\begin{aligned} &\int \text{div}(\nabla \Phi(\nabla \Psi(x) + tv) - x) \eta(\nabla \Psi) e^{-W} \\ &= \int (\text{Tr} D^2 \Phi(x + tv) \circ (\nabla \Psi) \cdot (I + tDv) \circ (\nabla \Psi) \cdot D^2 \Psi - d) \eta(\nabla \Psi) \, d\nu \\ &= \int (\text{Tr} D^2 \Phi(x + tv) \cdot (I + tDv) \cdot (D^2 \Phi(x))^{-1} - d) \eta \, d\mu. \end{aligned}$$

Remark that

$$\mathcal{T}(x, tv) = \text{Tr} D^2 \Phi(x + tv) \cdot (D^2 \Phi(x))^{-1} - d - \log[\det D^2 \Phi(x) \cdot (\det D^2 \Phi(x + tv))^{-1}] \geq 0.$$



Thus one obtains

$$\begin{aligned} \int (V(x+tv) - V(x))\eta \, d\mu &\geq \frac{K}{2} \int |\nabla\Phi(x+tv) - \nabla\Phi(x)|^2 \eta \, d\mu \\ &+ t \int \text{Tr}(D^2\Phi(x+tv) \cdot Dv(x) \cdot (D^2\Phi(x))^{-1})\eta \, d\mu \\ &+ \int \langle \nabla\Phi(x+tv) - \nabla\Phi(x), (D^2\Phi)\nabla\eta(\nabla\Phi(x)) \rangle \, d\mu + \int \mathcal{T}(x, tv)\eta \, d\mu. \end{aligned}$$

Now apply the same inequality to  $-tv$ , take the sum, and divide by  $t^2$ . It can be easily verified with the help of the Taylor formula that

$$\lim_{t \rightarrow 0} \frac{\mathcal{T}(x, tv) + \mathcal{T}(x, -tv)}{t^2} = \text{Tr}[(D^2\Phi)^{-1}(D^2\Phi)_v]^2 \geq 0.$$

In the limit  $t \rightarrow 0$  one gets the desired inequality.  $\square$

The proof of the Lemma 7.2 follows some elementary measure-theoretical arguments and we omit it here. It relies on the fact that the set of symmetric nonnegative matrices with multiple eigenvalue has smaller dimension in the ambient space of all symmetric nonnegative matrices.

**Lemma 7.2.** *Assume that  $\Phi$  is convex and twice continuously differentiable. For every  $\varepsilon > 0$  there exists a matrix  $Q_\varepsilon \geq 0$  such that  $\|Q_\varepsilon\| \leq \varepsilon$  and  $D^2\Phi + Q_\varepsilon$  has no multiple eigenvalues almost everywhere.*

**Theorem 7.3.** *Assume that  $D^2W \geq K \cdot \text{Id}$  and  $(D^2V)_+ \in L^1(\mu)$ . Then the following inequality holds*

$$\int \|(D^2V)^+\| \, d\mu \geq K \int \|D^2\Phi\|^2 \, d\mu.$$

*Proof. Step 1.* Let  $\Phi$  be smooth. Fix a point  $x_0$ . Assume that  $D^2\Phi(x_0)$  has no multiple eigenvalues. Assume that  $v$  is a smooth field coinciding with the unit eigenvectors of  $D^2\Phi$  corresponding to the unique largest eigenvalue  $\lambda$  in a neighborhood  $U_{x_0}$  of  $x_0$ . Let us show that  $\text{Tr}(\partial_v D^2\Phi \cdot Dv \cdot (D^2\Phi)^{-1}) \geq 0$  in  $U_{x_0}$ .

Indeed, one has

$$D^2\Phi \cdot v = \lambda \cdot v, \quad |v| = 1.$$

Differentiating both identities we get

$$(Dv)^T v = 0,$$

$$\partial_v D^2\Phi + D^2\Phi \cdot Dv = \lambda \cdot Dv + v \oplus \nabla\lambda.$$

Multiply (from the left) the second identity by  $(D^2\Phi)^{-1} \cdot (Dv)^T$  and take the trace. Taking into account that

$$\text{Tr}(D^2\Phi)^{-1}(Dv)^T \cdot v \oplus \nabla\lambda = \langle (D^2\Phi)^{-1}(Dv)^T v, \nabla\lambda \rangle = 0$$

one obtains

$$\text{Tr}(D^2\Phi)^{-1}(Dv)^T \cdot \partial_v D^2\Phi + \text{Tr}(D^2\Phi)^{-1}(Dv)^T D^2\Phi \cdot Dv = \lambda \cdot \text{Tr}(D^2\Phi)^{-1}(Dv)^T \cdot Dv.$$

Finally we get

$$\begin{aligned} \text{Tr}(\partial_v D^2\Phi \cdot Dv \cdot (D^2\Phi)^{-1}) &= \text{Tr}(D^2\Phi)^{-1}(Dv)^T \cdot \partial_v D^2\Phi \\ &= \text{Tr}(D^2\Phi)^{-1}(Dv)^T (\lambda I - D^2\Phi) \cdot Dv \\ &= \text{Tr}(D^2\Phi)^{-1/2}(Dv)^T (\lambda I - D^2\Phi) \cdot Dv \cdot (D^2\Phi)^{-1/2}. \end{aligned}$$

Note that the latter is equal to

$$\text{Tr}(ABA^T), \text{ where } A = (D^2\Phi)^{-1/2}(Dv)^T(D^2\Phi)^{-1/2}, \quad B = \lambda D^2\Phi - (D^2\Phi)^2.$$

Since  $\lambda$  is the largest eigenvalue,  $B$  is symmetric and non-negative. This immediately implies that

$$\text{Tr}(\partial_v D^2\Phi \cdot Dv \cdot (D^2\Phi)^{-1}) \geq 0$$

In particular, if  $\text{supp}(\eta) \subset U_{x_0}$ , we obtain from the previous lemma

$$(24) \quad \begin{aligned} \int \|(D^2V)^+\|^2 \eta \, d\mu &\geq K \int \lambda^2 \cdot \eta \, d\mu + \int \langle (D^2\Phi)_v \cdot v, (D^2\Phi)^{-1} \nabla \eta \rangle \, d\mu \\ &+ \int \text{Tr}[(D^2\Phi)^{-1}(D^2\Phi)_v]^2 \eta \, d\mu. \end{aligned}$$

**Step 2.** Let us assume that  $\Phi$  is a convex polynom such that  $D^2\Phi$  has no multiple eigenvalues almost everywhere. Recall that the set  $S := S(D^2\Phi)$ , where  $D^2\Phi$  has multiple eigenvalues, is the zero set of the discriminant of  $D^2\Phi$ . Hence  $S$  is an algebraic variety. In particular, for  $\mathcal{H}^{d-1}$ -almost every point  $x \in \partial S$  the set  $S \cap B_r(x)$  is diffeomorphic to  $\mathbb{R}^{d-1}$  for sufficiently small  $r$  (see [3], Proposition 3.3.14). Let  $\mathbb{R}^d \setminus S = \cup D_i$ , where every  $D_i$  is a connected component of  $\mathbb{R}^d \setminus S$ . Clearly, one can choose a vector field of unit eigenvectors  $v$  corresponding to the largest eigenvalue of  $D^2\Phi$  such that  $v|_{D_i}$  is smooth for every  $D_i$ .

By a classical result of Ky Fan [12] the function  $A \rightarrow \Lambda(A)$ , where  $\Lambda(A)$  is the largest eigenvalue is convex on the set of symmetric matrices. This implies, in particular, that the function

$$\lambda(x) : x \rightarrow \Lambda(D^2\Phi(x))$$

has a directional derivative

$$\partial_e \lambda(x) = \lim_{t \rightarrow 0} \frac{\lambda(x + te) - \lambda(x)}{t}$$

for every  $x$  and every direction  $e$ . For every regular point  $x \in \partial D_i$  we define

$$\nabla_{D_i} \lambda = \sum_{i=1}^d \partial_{e_i} \lambda \cdot e_i,$$

where the basis  $\{e_i\}$  is chosen in such a way that  $x + te_i \in D_i$  for the small values of  $t$  and every  $i$ .

Note that

$$\partial_e \langle D^2\Phi \cdot v, v \rangle = \langle D^2\Phi_e \cdot v, v \rangle + 2 \langle D^2\Phi \partial_e v, v \rangle$$

inside of  $D_i$ . Since  $v$  is an unit eigenvalue of  $D^2\Phi$ ,  $\partial_e v$  is orthogonal to  $D^2\Phi v$  and one has

$$\partial_e \lambda(x) = \langle D^2\Phi_e \cdot v, v \rangle.$$

Let us fix a compact domain  $B$  with smooth boundary and apply (24) to  $\nu = I_{D_i \cap B}$ . More precisely, we choose a sequence of smooth test functions  $\{\eta_m\}$  with supports inside of  $D_i \cap B_R$  such that  $\eta_i \rightarrow I_{D_i \cap B}$ . One gets in the limit

$$(25) \quad \begin{aligned} \int_{D_i \cap B} \|(D^2V)^+\|^2 \, d\mu &\geq K \int_{D_i \cap B} \lambda^2 \, d\mu + \int_{D_i \cap B} \text{Tr}[(D^2\Phi)^{-1}(D^2\Phi)_v]^2 \, d\mu \\ &+ \int_{\partial D_i \cap B} \langle \nabla_{D_i} \lambda, (D^2\Phi)^{-1} n_{D_i} \rangle \, d\mu + \int_{D_i \cap \partial B} \langle \nabla_{D_i} \lambda, (D^2\Phi)^{-1} n_B \rangle \, d\mu, \end{aligned}$$

where  $n_B$  is the inward normal to  $\partial B$ .

Now take a regular point  $x \in \partial D_i$ . Clearly,  $x$  belongs to the border between two sets  $D_i$  and  $D_j$ ,  $j \neq i$  and the inward normal of  $\partial D_i$  can be computed in the following way

$$n_{ij} = \frac{\nabla_{D_i} \lambda - \nabla_{D_j} \lambda}{\|\nabla_{D_i} \lambda - \nabla_{D_j} \lambda\|}.$$

Taking the sum of (25) over  $i$  we get that the integral term over the boundary  $\cup_i \partial D_i \cap B$  takes the form

$$\sum_{i,j} \int_{\partial D_i \cap \partial D_j \cap B} \langle \nabla_{D_i} \lambda - \nabla_{D_j} \lambda, (D^2 \Phi)^{-1} n_{ij} \rangle d\mu$$

and it is obviously non-negative.

Taking the sum over  $i$  we get

$$\begin{aligned} \int_B \|(D^2 V)^+\|^2 d\mu &\geq K \int_B \lambda^2 d\mu + \int_B \text{Tr}[(D^2 \Phi)^{-1} (D^2 \Phi)_v]^2 d\mu \\ &+ \sum_i \int_{D_i \cap \partial B} \langle \nabla_{D_i} \lambda, (D^2 \Phi)^{-1} n_B \rangle d\mu. \end{aligned}$$

Fix a smooth compactly supported nonnegative test function  $\xi$ . Applying the coarea formula and the above estimate applied to the level sets of  $\xi$  one can easily get that

$$\begin{aligned} \int \|(D^2 V)^+\|^2 \xi d\mu &\geq K \int \lambda^2 \xi d\mu + \int \text{Tr}[(D^2 \Phi)^{-1} (D^2 \Phi)_v]^2 \xi d\mu \\ &+ \sum_i \int_{D_i} \langle (D^2 \Phi)_v \cdot v, (D^2 \Phi)^{-1} \nabla \xi \rangle d\mu. \end{aligned}$$

Applying the standard relations between the operator and Hilbert-Schmidt norms

$$\begin{aligned} \text{Tr}[(D^2 \Phi)^{-1} (D^2 \Phi)_v]^2 &= \|(D^2 \Phi)^{-1/2} (D^2 \Phi)_v (D^2 \Phi)^{-1/2}\|_{HS}^2 \\ &\geq \frac{\|(D^2 \Phi)_v\|_{HS}^2}{\|D^2 \Phi\|^2} \geq \frac{\|(D^2 \Phi)_v\|^2}{\|D^2 \Phi\|^2}. \end{aligned}$$

and the Cauchy inequality one finally gets

$$(26) \quad \int \|(D^2 V)^+\|^2 \xi d\mu + 4 \int \frac{|\nabla \xi|^2}{\xi} d\mu \geq K \int \lambda^2 \xi d\mu.$$

Choosing an appropriate sequence of compactly supported functions  $\{\xi_n\}$  such that  $\lim_n \xi_n = 1$  and  $\lim_n \int \frac{|\nabla \xi_n|^2}{\xi_n} d\mu = 0$  we get the claim.

**Step 3.** Here we prove the general case. In the same way as in Theorem 3.1 one can approximate  $V$  and  $W$  by smooth functions with at most quadratic growth. Hence, one can assume without loss of generality that  $\Phi$  is smooth. To apply the previous step we fix a compact set  $B$  and choose a sequence of polynomial functions  $\{\Phi_n\}$  such that  $\Phi_n \rightarrow \Phi$  on  $B$  locally uniformly with all the derivatives up to the fourth order (this can be done by a multidimensional version of the Weierstrass approximation theorem).

Since we have convergence of the second derivatives, the functions  $\Phi_n$  are convex for sufficiently big  $n$ . Applying Lemma 7.2 we may assume that  $S(\Phi_n)$  has zero

measure. Note that the mapping  $\nabla\Phi_n$  sends  $e^{-V_n} dx$  onto  $\mu$ , where

$$V_n = W(\nabla\Phi_n) - \log \det D^2\Phi_n.$$

From the convergence  $\Phi_n \rightarrow \Phi$  follows that  $V_n \rightarrow V$  uniformly in  $B$  and the same holds for the derivatives up to the second order. Passing to the limits one obtains (26) for  $V$  and any smooth test function  $\xi$ . Choosing an appropriate sequence  $\{\xi_n\}$  with  $\xi_n \rightarrow 1$  one can easily complete the proof.  $\square$

The following result generalizes Theorem 7.3 in the same manner as Theorem 6.1 generalizes Theorem 3.1. The proof can be obtained by modifying the proof of Theorem 7.3 and we omit it here.

**Theorem 7.4.** *Assume that  $D^2W \geq K \cdot \text{Id}$ . Then for every  $r \geq 1$  one has*

$$K \left( \int \|D^2\Phi\|^{2r} d\mu \right)^{\frac{1}{r}} \leq \left( \int \|(D^2V)_+\|^r d\mu \right)^{\frac{1}{r}}.$$

This work was supported by the RFBR projects 07-01-00536 and 08-01-90431-Ukr, the DAAD Grant (2010), and the SFB701 at the University of Bielefeld.

#### REFERENCES

- [1] Ambrosio L., Gigli N., Savaré G., Gradient flows in metric spaces and in the Wasserstein spaces of probability measures., Lectures in Math., ETH Zurich, 2005.
- [2] Bakelman I.J., Convex analysis and nonlinear geometric elliptic equations., Springer-Verlag, Berlin, 1994.
- [3] Bochnak J., Coste M., Roy M.-F., Real Algebraic Geometry, Springer Verlag, Berlin-Heidelberg-New York, 1998.
- [4] Bogachev V.I., Measure Theory. Springer-Verlag, Berlin, 2007.
- [5] Bogachev V.I., Kolesnikov A.V., Integrability of absolutely continuous transformations of measures and applications to optimal mass transport. Probab. theory and its appl., 50(3), (2005), 3–25.
- [6] Caffarelli L.A., Interior  $W^{2,p}$ -estimates for solutions of the Monge-Ampère equation. Ann. of Math. (2), 131(1), (1990), 135–150.
- [7] Caffarelli L.A., Monotonicity properties of optimal transportation and the FKG and related inequalities, Comm. Math. Phys. 214(3), (2000), 547–563.
- [8] Caffarelli L.A., Cabré X., Fully nonlinear elliptic equations, Amer. Math. Soc., Providence, RI, 1995.
- [9] Carlen E.A., Superadditivity of Fisher’s information and logarithmic Sobolev inequalities, J. Funct. Anal., 101 (1991), 194–211.
- [10] Delanoë P., Loeper G., Gradient estimates for potentials of invertible gradient mappings on the sphere. Calc. Var. Partial Differential Equations, 26(3), (2006), 297–311.
- [11] Eberle A., Uniqueness and non-uniqueness of semigroups generated by singular diffusion operators. Springer, Lecture Notes in Math., (1999), 1718.
- [12] Fan K., On a theorem of Weyl concerning eigenvalues of linear transformations. I., Proc. Nat. Acad. Sci. U. S. A., 35 (1949), 652–655.
- [13] Figalli A., Kim Y.-H., McCann R.J., Regularity of optimal transport maps on multiple products of spheres. arxiv: 1006.1957.
- [14] Gilbarg D., Trudinger N.S., Elliptic partial differential equation of the second order, Springer-Verlag, Berlin, 2001.
- [15] Gozlan N., Leonard Ch., Transport inequalities. A survey. Markov Processes and Related Fields, 16, 2010, 635–736
- [16] Gutiérrez C.E., The Monge-Ampère equation, Progress in Nonlinear Differential Equations and Their Applications, 44, Birkhäuser, 2001.
- [17] Ivchikina N.M., Construction of a priori bounds for convex solutions of the Monge-Ampère equation by integral methods, Ukrian. Math. 30 (1978), 32–38.
- [18] Kim Y.-H., Milman E., A Generalization of Caffarelli’s Contraction Theorem via (reverse) Heat Flow, arXiv:1002.0373.

- [19] Kim Y.-H., McCann R.-J., Continuity, curvature, and the general covariance of optimal transportation. *J. Eur. Math. Soc. (JEMS)* 12 (2010) 1009-1040
- [20] Kolesnikov A.V., Convexity inequalities and optimal transport of infinite-dimensional measures. *J. Math. Pures Appl.* 83 (11) (2004), 1373–1404.
- [21] Kolesnikov A.V., On global Hölder estimates for optimal transportation. *Mat. zametki*, 88:5 (2010), 708-728.
- [22] Kolesnikov A.V., Röckner M. (work in progress).
- [23] Krylov N.V., Fully nonlinear second order elliptic equations: recent developments. *Ann. Scuola Norm. Sup. Pisa Cl. Sci* (4), Vol. XXV (1997), 569–595.
- [24] Ledoux M., The concentration of measure phenomenon, *Mathematical Surveys and Monographs* 89. Amer. Math. Soc., 2001.
- [25] Loeper G., On the regularity of solutions of optimal transportation problems. *Acta Math.* 202 (2), (2009), 241–283.
- [26] McCann R.J., A convexity principle for interacting gases. *Adv. Math.*, 128 (1997), 153-179.
- [27] Milman E., On the role of convexity in functional and isoperimetric inequalities, *Proc. London Math. Soc.* 99 (3), 32–66, 2009.
- [28] Milman E., On the role of convexity in isoperimetry, spectral gap and concentration, *Invent. Math.* 177 (1), 1–43, 2009.
- [29] Ovsienko Yu.V., Zhdanov R.I., Estimates for Sobolev norm of triangular mappings, *Moscow Univ. Math. Bull.*, 62(1), (2007), 1–4.
- [30] Pogorelov A.V., *Monge-Ampère equations of elliptic type*, Noordhoff, 1964.
- [31] Talagrand M., Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6, (1996), 587–600.
- [32] Trudinger N.S., X.-J. Wang, On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) 8 (1), (2009), 143–174.
- [33] Valdimarsson S. I., On the Hessian of optimal transport potential, *Ann. Sc. Norm. Sup. Pisa Cl. Sci.* (5), 6(3), (2007), 441–456.
- [34] Villani C., *Topics in Optimal Transportation*, Amer. Math. Soc. Providence, Rhode Island, 2003.